We would like to thank the reviewers and editor for the very useful comments and suggestions. We reviewed the text to incorporate all the comments and added a table to provide the readers with references to main international integrators.

Final edit 08/03/25 – we corrected the indicated link

Answer to Rev 1.

in the following the reviewed text (without figures and tables):

1 Introduction

The importance of ocean observation in met-ocean forecasting is emphasized, as it provides crucial data for understanding oceanic behaviour and coastal areas. The integration of parameters like temperature, salinity, currents, and atmospheric conditions enhances model accuracy, crucial for effective management of human impacts and resource exploitation. The complex ocean data collection framework involves numerous in situ platforms (Figure 1), remote sensors, and types of data, necessitating the provision of multidisciplinary, aggregated datasets (Belbéoch et al., 2022). Marine data integrators play a pivotal role in managing, integrating, and advancing the understanding of marine environments. They collect, process, and analyse diverse data types to create comprehensive datasets, contributing to informed decision-making in areas such as fisheries management, offshore energy development, marine conservation, etc (see e.g, Novellino et al., 2024). Additionally, these integrators support the development of technologies for monitoring the marine environment, continually refining data collection processes to enhance accuracy. Over the past three decades, progress in marine data management has been marked by the establishment of international programs and networks, such as the International Oceanographic Data and Information Exchange (IODE), the Global Ocean Observing System (GOOS), the IOC Ocean Data and Information System (ODIS). These initiatives, including the World Ocean Database, involve collaborative efforts globally, led by organizations like the Intergovernmental Oceanographic Commission (IOC), the World Meteorological Organization (WMO), the Environment Program of the United Nations (UNEP), the International Council for the Exploration of the Sea (ICES), etc. Under the GOOS framework (Figure 2), the Observation Coordination Group (OCG), supported by OceanOPS (the GOOS in-situ Ocean Observations Programme Support Centre) and GOOS Regional Alliances (GRAs), coordinates the GOOS observing networks to provide ocean observing information (Moltmann et al., 2019). GRAs integrate national monitoring needs into a regional system, facilitating data assembly and exchange (Corredor, 2018). Data Assembly Centres (DACs) and Global DACs (GDACs) play a critical role in this process by receiving, quality-controlling, and assembling data from various sources. They act as primary access points for this information, adhering to a common data format (netCDF). Despite these efforts, GOOS networks and data represent only a subset of the overall ocean data framework. While progress has been made in modernizing the WMO data exchange system—transitioning from the Global Telecommunication System (GTS) to WIS 2.0—by leveraging new web technologies and existing DAC/GDAC infrastructures, full data integration between OCG networks and national/regional initiatives has yet to be achieved. In this intricate and dispersed framework, integration services play a crucial role in harmonizing metadata, applying standardized data quality checks, and facilitating the integration of diverse datasets and models. GOOS networks, guided by the OCG data strategy (O'Brien et al., 2024), are establishing global data nodes that progressively enhance overall data delivery while maintaining "GOOS quality" within the broader ocean data lake. Furthermore, the adoption of unified controlled vocabularies, common data models, and standardized transport formats ensures the seamless integration of real-time, near real-time (NRT), and delayed-mode (DM) observations into numerical models. The following section describes the European marine data integration landscape which is characterized by three prominent initiatives: the Copernicus Marine Service (the In-Situ Thematic Assembly Centre), the European Marine Observation and Data network (with a focus on Physics) and the SeaDataNet network of the National Oceanographic Data Centres associated with the International Oceanographic Commission.

2 European Marine Data Integrators

To exemplify the importance of data integrators, a few relevant examples from Europe are presented.

Copernicus Marine In-Situ Thematic Assembly Centre (Copernicus Marine INS TAC). Within this programme, the Copernicus INS TAC is a distributed service integrating data from different sources for operational needs in oceanography. The Copernicus INS TAC integrates and quality controls in a homogeneous manner in situ data from data providers in order to fit the needs of internal and external users. It provides access to integrated datasets of core parameters for initialization, assimilation into and validation of ocean numerical models, which are used for forecasting, analysis and re-analysis of ocean physical and biogeochemical conditions. Since the primary objective of Copernicus Marine is to forecast ocean state, the initial focus has been on observations from autonomous observatories at sea (e.g. floats, buoys, gliders, ferryboxes, drifters, and ships of opportunity). The second objective is to set up a system for re-analysis purposes that requires products integrated over the past 25 to 60 years. The Copernicus Marine INS TAC comprises a global in-situ centre and 6 regional in-situ centres, one for each EuroGOOS Regional Ocean Observing System (ROOS). The INS TAC has been designed to fulfil the Copernicus Marine Core Service needs and the EuroGOOS ROOS needs. The focus is on essential ocean variables (EOVs) that are presently necessary for Copernicus Monitoring and Forecasting Centres, namely temperature, salinity, sea level, current, waves, chlorophyll / fluorescence, oxygen and nutrients. Additional atmospheric parameters (such as wind, air temperature, air pressure, etc.) are added by some ROOSes to these regional in-situ portals to fulfil additional downstream applications needs. For near-real-time and delayed mode products, the Copernicus Marine In-Situ Thematic Assembly Centre is connected to the GOOS global networks and each Regional Ocean Observing System (ROOS) of EuroGOOS. In the case of DM products, it is also connected to the SeaDataNet Network, which comprises National Oceanographic Data Centres (NODCs). The Copernicus INS TAC integrates data from various observation programs, including Argo, OceanGliders, Data Buoy Cooperation Panel (DBCP), OceanSITES, and ship data obtained via NODCs, leveraging the AtlantOS observations. Whenever possible, the Copernicus INS TAC adheres to the standards developed within the SeaDataNet framework.

European Marine Observation and Data Network (EMODnet). The European Marine Observation and Data Network (EMODnet) is the EU infrastructure for in situ marine data. The goal of EMODnet is to provide access to a wide range of standardized and harmonized marine data, making it easier for researchers, policymakers, and the public to access and use marine information. EMODnet focuses on various thematic areas, including bathymetry, geology, physics, chemistry, biology, and human activities in the marine environment (Shepherd, 2018). By pooling and harmonizing data from various sources, EMODnet aims to create a comprehensive and easily accessible marine data infrastructure that supports a wide range of marine and maritime activities (Schaap et al, 2022). EMODnet Physics (https://emodnet.ec.europa.eu/en/physics, Figure 3) is the domain-specific project (Míguez et al., 2019) that provides in situ ocean physics data and data products built with common standards, free of charge, and without restrictions. These services encompass a wide range of parameters, including temperature, salinity, current profiles, sea level trends, wave height and period, wind speed and direction, water turbidity (light attenuation), underwater noise, river flow, and sea-ice coverage. EMODnet Physics offers an array of in situ data collections (time-series, profiles, and datasets) obtained from various platforms (such as tide gauges, river stations, floats, buoys, gliders, drifters, and ship-based observations). EMODnet Physics does not operate platforms; instead, it integrates and federates key data infrastructures and programs. For example, it is synchronized with Copernicus Marine INS TAC and includes supplementary in situ data from PANGAEA (https://www.pangaea.de/), the International Council for the Exploration of the Sea (https://www.ices.dk/data/data-portals/Pages/ocean.aspx), the European Multidisciplinary Seafloor and Water Column Observatory (EMSO) (https://emso.eu/), the SeaDataNet network of National Oceanographic Data Centres (NODCs), and other Global Ocean Observation System networks (https://goosocean.org/). The data and data products are accompanied by metadata, offering users comprehensive information regarding the provenance, content, location, time, data sources, and quality check procedures. It supports human-based data discovery (https://https://emodnet.ec.europa.eu/geoviewer/) and machine-to-machine interoperability (https://data-erddap.emodnet-physics.eu/erddap/) and contributes to enhancing our understanding of the physical aspects of the marine environment. EMODnet Physics supports various applications, including scientific research, coastal management, maritime operations, and policymaking.

SeaDataNet. SeaDataNet (http://www.seadatanet.org) is a Pan-European network of professional marine data centres providing data and metadata standards for the marine community, and on-line access to their data holdings of standardized quality (Schaap and Lowry, 2010). Founding partners are the National Oceanographic Data Centres, major marine research institutes, UNESCO-IOC, ICES, and European Commission Joint Research Centre (EC JRC). Over three decades, SeaDataNet has expanded its network of data centres and infrastructure in a long series of EU projects, mostly funded through EU DG RTD. SeaDataNet operates an infrastructure for managing, indexing and providing access to ocean and marine environmental data sets and data products (e.g. physical, chemical, geological, and biological properties) and for safeguarding the long term archival and stewardship of these data sets. Data are derived from many different sensors installed on research vessels, satellites and in-situ platforms that are part of various ocean and marine observing systems and research programs. A core SeaDataNet service is the Common Data Index (CDI) data discovery and access service which provides harmonized discovery and access to a large volume of marine and ocean data sets. Currently, more than 110 data centres are connected to the CDI service from 34 countries around European seas, giving access to more than 2.5 million data sets, originating from more than 650 organizations in Europe. This imposes strong requirements towards ensuring quality, elimination of duplicate data and overall coherence of the integrated data set. This is achieved in SeaDataNet by establishing and maintaining accurate metadata directories and data access services, as well as common standards like vocabularies, metadata formats, data exchange formats, quality control methods and quality flags. SeaDataNet data resources are quality controlled and are major input for developing added-value services and products that serve users from government, research and industry (Simoncelli et al., 2022).

3 Single Source Integrators

Besides these key European multi-parameter ocean data integrators, there are a number of initiatives that focus on single platforms or specific ocean variables. These initiatives concentrate on specific aspects of the marine environment, targeting a particular platform or variable for data collection and integration. Examples include projects that solely focus on buoys or floats for collecting oceanic data, or initiatives that specifically address parameters such as sea surface temperature, ocean currents, or marine biodiversity. By specializing in a single platform or variable, they can provide detailed and focused data products and services that cater to specific user needs and applications, as well as provide a simplified source for specific forecasting systems. The following Table 1 summarizes the most used ones.

4 Ways forward in ocean data integration

In advancing ocean data integration, several key strategies can push our understanding of marine ecosystems and facilitate more informed decision-making. Shared data repositories and standardized data formats can streamline the integration process, ensuring compatibility and accessibility, and more generically data-FAIR (Wilkinson et al. 2016). Harnessing the power of emerging technologies, such as artificial intelligence and machine learning, offers opportunities to analyse vast datasets swiftly and extract meaningful insights. Implementing autonomous sensors and advanced monitoring systems enhances real-time data collection, providing a more comprehensive and dynamic picture of oceanic conditions. To follow the evolution of ocean general metocean models in terms of spatial resolution, which, in the future, will reach the kilometric scale at the global level, there is a clear need for more sensors deployed at the global, regional, and local scale. In this framework, the inclusion of cost-effective and citizen-based data collection is also a key forward-looking and long-term initiatives, such EMODnet, may have a crucial role in setting up the data flow capacities for emerging networks not organized under GOOS networks. Timeliness is also an important parameter to be improved to ensure that data are available at each model run, particularly crucial for coastal applications where ocean dynamics evolve rapidly. Nevertheless, data usability/consumability strongly depends on the data policy license, and there is an increasing push for adopting the Common Creative framework and, in particular, the CC-BY license, where the only limitation is that credit must be given to the creator. Integrating these strategies collectively will not only advance ocean data integration but also contribute to the ongoing evolution of ocean general metocean models, including digital twins of the oceans, and foster a more comprehensive and accessible understanding of the marine environment.

Answer to rev. 2

We are reshaping the paper to incorporate as much feedback as possible while maintaining brevity and conciseness.

Introduction was reorganized as follow:

The importance of ocean observation in met-ocean forecasting is emphasized, as it provides crucial data for understanding oceanic behaviour and coastal areas. The integration of parameters like temperature, salinity, currents, and atmospheric conditions enhances model accuracy, crucial for effective management of human impacts and resource exploitation. The complex ocean data collection framework involves numerous in situ platforms (Figure 1), remote sensors, and types of data, necessitating the provision of multidisciplinary, aggregated datasets (Belbéoch et al., 2022). Marine data integrators play a pivotal role in managing, integrating, and advancing the understanding of marine environments. They collect, process, and analyse diverse data types to create comprehensive datasets, contributing to informed decision-making in areas such as fisheries management, offshore energy development, marine conservation, etc (see e.g, Novellino et al., 2024). Additionally, these integrators support the development of technologies for monitoring the marine environment, continually refining data collection processes to enhance accuracy. Over the past three decades, progress in marine data management has been marked by the establishment of international programs and networks, such as the International Oceanographic Data and Information Exchange (IODE), the Global Ocean Observing System (GOOS), the IOC Ocean Data and Information System (ODIS). These initiatives, including the World Ocean Database, involve collaborative efforts globally, led by organizations like the Intergovernmental Oceanographic Commission (IOC), the World Meteorological Organization (WMO), the Environment Program of the United Nations (UNEP), the International Council for the Exploration of the Sea (ICES), etc. Under the GOOS framework (Figure 2), the Observation Coordination Group (OCG), supported by OceanOPS (the GOOS in-situ Ocean Observations Programme Support Centre) and GOOS Regional Alliances (GRAs), coordinates the GOOS observing networks to provide ocean observing information (Moltmann et al., 2019). GRAs integrate national monitoring needs into a regional system, facilitating data assembly and exchange (Corredor, 2018). Data Assembly Centres (DACs) and Global DACs (GDACs) play a critical role in this process by receiving, quality-controlling, and assembling data from various sources. They act as primary access points for this information, adhering to a common data format (netCDF). Despite these efforts, GOOS networks and data represent only a subset of the overall ocean data framework. While progress has been made in modernizing the WMO data exchange system—transitioning from the Global Telecommunication System (GTS) to WIS 2.0—by leveraging new web technologies and existing DAC/GDAC infrastructures, full data integration between OCG networks and national/regional initiatives has yet to be achieved. In this intricate and dispersed framework, integration services play a crucial role in harmonizing metadata, applying standardized data quality checks, and facilitating the integration of diverse datasets and models. GOOS networks, guided by the OCG data strategy (O'Brien et al., 2024), are establishing global data nodes that progressively enhance overall data delivery while maintaining "GOOS quality" within the broader ocean data lake. Furthermore, the adoption of unified controlled vocabularies, common data models, and standardized transport formats ensures the seamless integration of real-time, near real-time (NRT), and delayed-mode (DM) observations into numerical models.

We also included a new figure (about GOOS framework) and we are considering to add a clear diagram to map the initiatives.

We also reviewed the final paragraph as follows:

In advancing ocean data integration, several key strategies can push our understanding of marine ecosystems and facilitate more informed decision-making. Shared data repositories and standardized data formats can streamline the integration process, ensuring compatibility and accessibility, and more generically data-FAIR (Wilkinson et al. 2016). Harnessing the power of emerging technologies, such as artificial intelligence and machine learning, offers opportunities to analyse vast datasets swiftly and extract meaningful insights. Implementing autonomous sensors and advanced monitoring systems enhances real-time data collection, providing a more comprehensive and dynamic picture of oceanic conditions. To follow the evolution of ocean general metocean models in terms of spatial resolution, which, in the future, will reach the kilometric scale at the global level, there is a clear need for more sensors deployed at the global, regional, and local scale. In this framework, the inclusion of cost-effective and citizen-based data collection is also a key forward-looking and long-term initiatives, such EMODnet, may have a crucial role in setting up the data flow capacities for emerging networks not organized under GOOS networks. Timeliness is also an important parameter to be improved to ensure that data are available at each model run, particularly crucial for coastal applications where ocean dynamics evolve rapidly. Nevertheless, data usability/consumability strongly depends on the data policy license, and there is an increasing push for adopting the Common Creative framework and, in particular, the CC-BY license, where the only limitation is that credit must be given to the creator. Integrating these strategies collectively will not only advance ocean data integration but also contribute to the ongoing evolution of ocean general metocean models, including digital twins of the oceans, and foster a more comprehensive and accessible understanding of the marine environment.